

A Theory of Scientific Study

Robert W.P. Luk

Department of Computing, The Hong Kong Polytechnic University

Hung Hom, Kowloon, Hong Kong

csrluk@comp.polyu.edu.hk

This paper presents a theory of scientific study which is regarded as a social learning process of (working) scientific knowledge creation, revision, application, monitoring (e.g., confirmation) and dissemination (e.g., publication) with the aim of securing good quality, general, objective, testable and complete scientific knowledge of the domain. The theory stipulates the aim of scientific study that forms the basis of its principles. It also makes seven assumptions about scientific study and defines the major participating entities (i.e., scientists, scientific knowledge and enabling technical knowledge). It extends a recent process model of scientific study into a detailed interaction model as this process model already addresses many issues of philosophy of science. The detailed interaction model of scientific study provides a common template of scientific activities for developing logical (data) models in different scientific disciplines (for physical database implementation), or alternatively for developing (domain) ontologies of different scientific disciplines. Differences between research and scientific studies are discussed, and a possible way to develop a scientific theory of scientific study is described.

Keywords: *Science, Scientific Study, Scientific Inquiry, Theory and Model*

1. Introduction

The scientific method (e.g., (Pierce, 1878; Weston, 1987)) has received a great deal of attention and enjoyed a lot of empirical success. However, its validity as the only accepted method of scientific inquiry has been questioned (e.g., (Cleland, 2001)). Recently, it is thought to be only one type of activity in scientific study by Luk (2010) who distinguishes scientific study from other types of study (e.g., criminal investigation) not just by the process but also by its knowledge elements and by their roles. Using the study by Luk (2010), this paper presents a theory of scientific study. In doing so, the paper also borrows heavily his terminology.

We are motivated to develop a theory of scientific study that forms the basis to conduct scientific study. This basis distills scientific studies of different (scientific) disciplines into the common principles, differentiates the basic principles from the guiding ones, reveals the underlying assumptions and shows how these are related to the aim of scientific study. This aim reflects the desired qualities that scientific knowledge should possess in any (scientific) discipline. To reach these qualities, scientific study is seen as a social learning process that guides the creation, the revision, the application, the monitoring and the dissemination of scientific knowledge towards achieving those desired qualities (e.g., accuracy, reliability and objectivity).

We claim that it is possible to develop a theory of scientific study by construction. Apart from the principles and assumptions, this theory has a general (contextual) model of scientific study as a social learning process. Focusing on carrying out the scientific study, a more detailed model integrates the process model by Luk (2010) with the scientist entity, scientific knowledge entity and enabling technical knowledge entity. At greater details, the theory entity, scientific model entity and the experiment entity are seen as entity clusters, the details of which are presented and discussed in turn. An

integrated model of these entity clusters serves as a template for the practicing scientists to develop their model of scientific study of their (scientific) disciplines while maintaining some common understanding of the scientific study across scientific disciplines by using the (common) template. To carry out scientific study, we claim that scientists know enabling technical knowledge, some of which is more domain independent (e.g., mathematics) and some is domain-specific (e.g., Feynman diagram).

To support our claim, the general models and the detailed ones are presented for discussion. For wider acceptance, our general model is based on the process model of scientific study by Luk (2010), which is argued to be able to differentiate from other studies, and which tries to integrate the different issues in philosophy of science by putting them into the context of scientific study. To support the claim that scientists possess enabling technical knowledge, the figure in Section 4.5 lists examples of domain specific/domain independent enabling technical knowledge, and it shows how such knowledge interrelates with scientific knowledge.

The significance of this work is that it invites other theories of scientific study to be developed by improving or by developing a better theory than our initial theory. This work may lead to the construction of a scientific theory of scientific study. It also provides a (common) template for structuring scientific knowledge. More specifically, it is designed to be further developed into detailed data models of scientific study in specific (scientific) disciplines for their knowledge management needs (e.g., (Kingston, 2002)), and it can naturally integrate recent work on ontologies (e.g., in geosciences (Bordaric and Gahegan, 2006), in economics (Pratten, 2007) and in scientific experiments (Soldatova and King, 2006)), on the detailed scientific workflows (e.g., (Ludäscher et al., 2006)) and on scientific infrastructure (e.g., (Hars, 2001)). Such integration complements current e-science activities (e.g., De Roure et al., 2003).

The rest of this paper is organized as follows. Section two describes a contextual interaction model of scientific study as a social learning process. As this is the most general level interaction model which is encountered first, this section is also used as an introduction to interaction models. Section three sets up our theory of scientific study. This theory stipulates the aim, principles and assumptions of scientific studies, as well as the defining characteristics of scientists. Section four extends the simplified model of scientific study in section two to our initial, detailed (interaction) model of scientific study so that our theory of scientific study can be applied to this interaction model (similar to the scientific study, physics, where Newton's laws of motion (i.e., the theory) are applied to build mechanical models as in Luk (2010)). This section also shows an interaction model of the different types of knowledge possessed by scientists. Section five discusses the differences between research and scientific study. Section six describes a possible way to develop a scientific theory of scientific study. Section seven is the related work. Finally, the paper concludes.

We follow the definitions of the following terms in (Luk, 2010): science, scientific knowledge, scientific study, scientific research, research, theory, scientific theory, (computational/conventional) scientific model, experiment, physical situation, formative scientific study, developing scientific study and mature scientific study. Furthermore, the term "scientific study" in this paper means doing an investigation as an activity in a formative, or developing or mature scientific study (Luk, 2010). When we refer to a domain of study in science, we use terms like "domain of study" or "(scientific) discipline", in order to avoid misreading scientific studies as (scientific) disciplines.

2. Contextual Interaction Model of Scientific Study

The contextual interaction model of scientific study has the highest level, participating entities which interact with the scientific study. The basic constructs of a contextual interaction model are:

- (a) *internal entities* (e.g., scientific study in Figure 1) are entities or components of scientific study. These are presented by rectangles;
- (b) *external entities* (e.g., scientist in Figure 1) are the highest level, participating entities that are interacting with the scientific study entity. External entities are represented by rectangles with rounded corners;
- (c) *relationship* between entities is represented by a line. The name of the relationship is the label with an arrow next to the relationship. For example, the two relationships between the scientific study entity and the physical situation entity are the excite relationship and the measured_by relationship (Figure 1). The arrow next to the name of the relationship shows the direction of reading the relation between the participating entities and the relationship. For example, the excite relationship can be read as “scientific study excites the physical situation”. Likewise, the measured_by relationship is read as “the physical situation is measured by scientific study”.

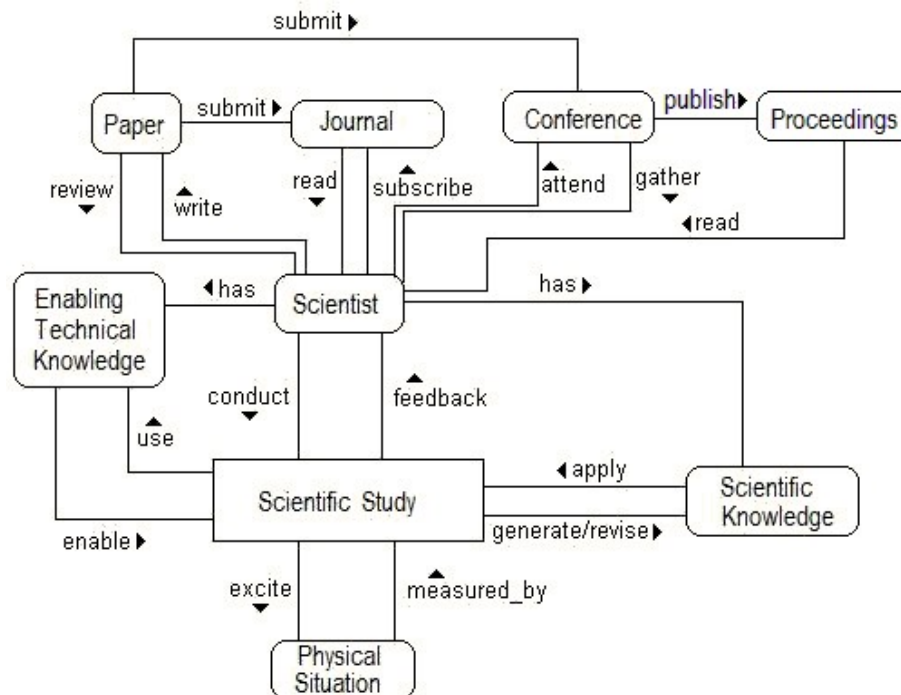


Figure 1: A contextual interaction model of scientific study.

An interaction model is a certain detailed part of a contextual interaction model. A (contextual) interaction model is almost the same as the entity-relationship (ER) diagram (Hoffer et al., 2002) in conceptual modeling of databases except that

- (a) the attributes of the entities are not shown in the interaction model for clarity of presentation; and
- (b) the cardinality constraints are not shown in the interaction model for clarity of presentation.

If the attributes and the cardinality constraints are added to the interaction model, it will become an ER model, which can be converted into a logical model (Hoffer et al., 2002; Silberschatz et al., 2005), in turn. Such a logical model has logical implications or rules (called functional dependencies) for checking data consistency. This logical model can be implemented as a physical model (Hoffer et al., 2002) that exists in a database which keeps track of the data. The logical model can be considered as a component of the scientific model of scientific study, and its physical model as a component of the computational scientific model of scientific study.

Instead of developing into an ER model or an enhanced ER (EER) model, the (contextual) interaction model can be developed into ontologies. The entities can be treated as concepts, and the relationships can be treated as relations between concepts. While the ontology can be developed from the interaction model, we prefer to develop the interaction model using some EER constructs (like entity cluster) in the rest of this paper because of the following reasons. First, an ontology typically develops its concepts vertically using *is-a* relations and *part-of* relations showing greater and greater details of the model (see Soldatova and King, 2006 for example). In this paper, we do not want to develop the interaction model vertically. Instead, we want to develop the interaction model horizontally reaching different aspects of scientific study so that the reader can have an overview picture of scientific study. Second, we want to hide the detailed low-level concepts (of the ontologies) because such details (e.g., experiment goal concept in Figure 1 of Soldatova and King, 2006) are irrelevant to philosophy of science. Third, we want to develop the interaction model horizontally in order to show the different relationships between different entities whereas ontologies typically discuss only *is-a* or *part-of* relations without the other relations in the model. Fourth, it is clearer to show the interaction model using some EER constructs instead of ontologies because we develop the interaction model horizontally without the *part-of* relations, axioms, rules and detailed (low-level) concepts which would otherwise clutter up the diagrams. Fifth, our interaction model has recursive relationship as we show later, which is allowed by EER diagrams, but it is not clear whether this kind of relationship is allowed in ontologies. Sixth, apart from the recursive relationship, it is possible to develop the interaction model using some EER constructs into ontologies, so we do not lose the generality of the interaction model (i.e., it can be converted into EER model or ontologies). While we show how the interaction model can be developed into EER models and then data models, etc., the reader can ignore this and readily develop the interaction model into ontologies if desired. Seventh, EER diagrams have *is-a*

relationships which can easily map to *is-a* relations in ontologies, so our interaction model with EER constructs does not lack this kind of expressiveness of ontologies. Eighth, the interaction model using EER constructs can readily integrate with the database system for journal/conference review processes which are captured by database technology possibly using EER constructs (see Figure 1 for example). Therefore, in the rest of this paper, we will discuss the interaction model as if though we want to develop the interaction model into EER models and let the reader to develop the interaction model into ontologies for himself/herself if desired.

The contextual interaction model in Figure 1 illustrates the social learning process known as scientific study. This social learning process serves to create, revise, apply, monitor and disseminate the (working) scientific knowledge which consists of theories, models and experiments. Through this process, scientists conduct scientific studies to obtain their feedback from the physical situation. The scientists should possess both scientific knowledge and enabling technical knowledge. The scientific knowledge is the knowledge about the particular domain of science, and it is put in the form of theories, scientific models, etc (as in Luk, 2010). The enabling technical knowledge is also possessed by the scientists for them to conduct the scientific studies in general. Some enabling technical knowledge facilitates the scientists not only to conduct scientific study in the specific domain but in general any domain. Examples of such enabling technical knowledge are mathematics, logic, research methodologies, etc. During scientific study, certain scientific knowledge is applied and the scientific study eventually generates new scientific knowledge or revisions of existing scientific knowledge.

The contextual interaction model in Figure 1 also shows that the scientists engage in publications and conferences. These social practices are important to assure the objectivity of scientific studies by disclosing the scientific studies

and their findings to peer scientists for evaluation, confirmation and further advancement of the domain. In fact, this part of the contextual interaction model about journal publications and conference activities are increasingly being managed by using database technology (e.g., ManuscriptCentral for the ACM and IEEE academic societies, EES for Elsevier and EditorialManager for Springer), so such database technology already captures some part of the contextual interaction model of scientific study. Note that the dissemination of (working) scientific knowledge to the public is not limited by journal publication or conference proceedings. Therefore, Figure 1 serves only to illustrate the dissemination rather than providing an exhaustive list of dissemination (which may include book publication, Internet publication, etc).

There are social processes (e.g., lobbying journal editors) of scientists other than dissemination (as in Figure 1) but these processes are not considered to have general significance like the social process of dissemination (which may include publishing in magazines, in newspapers, or through the Internet). For example, one important social process mentioned by Latour (1987) is acquiring research funding. For some research, it may even prevent the research to be done. However, not every piece of research requires funding especially the paper-and-pencil type of research. For example, Albert Einstein did not acquire any research funding to carry out the work on special relativity. So, such social process does not have general significance for scientific study, and it is not included in our theory of scientific study. However, the social process of dissemination is important for the objectivity of science, for the verification of theories, for extending existing work, etc. So, this social process has general significance for scientific study and it is included in the contextual part of our theory.

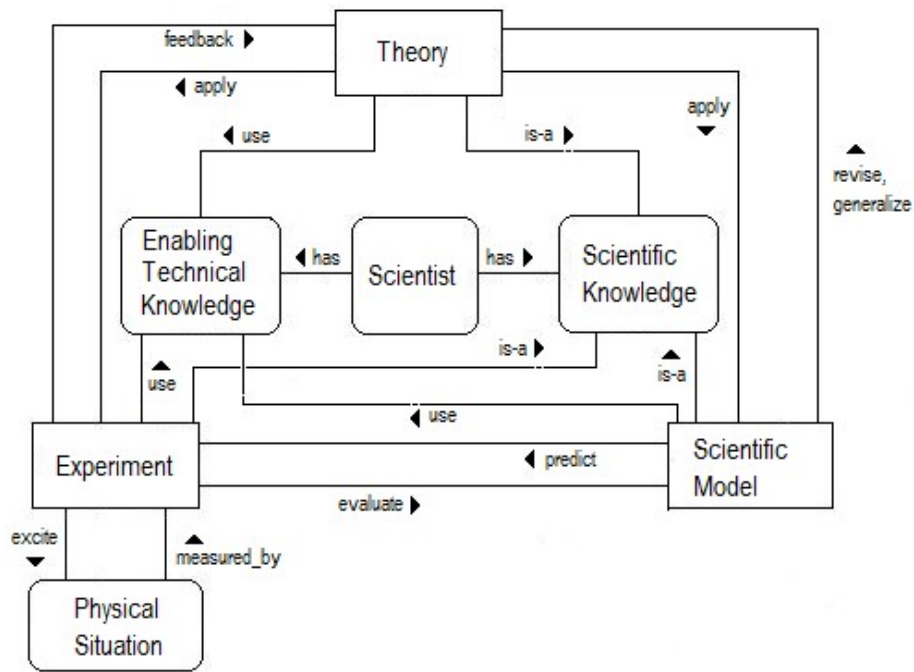


Figure 2: A simplified interaction model of scientific study based on the contextual interaction model in Figure 1 and the process model in Figure 1 of (Luk, 2010).

In Figure 1, scientific study is shown as an entity. Actually, it can be regarded as an entity cluster that contains theory entity, scientific model entity and experiment entity as shown in Figure 2. Essentially, we regard scientific study as the (social) learning process of applying theories to build scientific models or using theories to explain phenomena, where the experiment verifies the theory and evaluates the scientific models, based on measurements from the physical situations. This essentially corresponds to Figure 1 of Luk (2010). In here, Figure 2 adds the scientific knowledge entity, the enabling technical knowledge entity and the scientist entity into the diagram (Figure 1 of (Luk, 2010)) showing how these entities are related to the scientific study component entities: theory, scientific model and experiments. The simplified interaction model in Figure 2 illustrates that the scientific knowledge and enabling technical knowledge that exist outside the interaction model can be applied or specialized as the knowledge about the theory, the scientific model and the experiments needed for the scientific study. For example, the

scientific knowledge of a particular domain may have a number of theories but only one theory may be used in a particular scientific study. Having discussed the context in which a scientific study is conducted, the next section sets up a theory of scientific study.

3. An Initial Theory of Scientific Study

Our theory of scientific study consists of a set of general statements that are grouped into definitions, principles and assumptions. Many of these general statements, for instance principles and assumptions, are derived from issues in philosophy of science, and are related to its underlying aim. Our theory explains how scientific study should be conducted according to the mentioned principles after making its assumptions. It relies on the social process of (academic) publication to encourage scientific studies to be reliable, accurate, consistent, testable, objective and complete. It also relies on its aim, its basic principle of empiricism and its guiding principle of investigation objectivity to ensure the created knowledge is scientific (rather than, for instance, being purely mathematical and unrelated to science).

3.1 Aim and Definition

Our theory of scientific study asserts the following general statement:

Definition 1: The aim of *scientific study* (as an activity) is (i) to produce good quality, objective, general, testable and complete *scientific knowledge* of the chosen domain of study (called context), and (ii) to monitor/apply such knowledge.

The underlined terms in indented passages of this paper are defined elsewhere in this paper whereas those in italic in indented passages of this

paper are defined in (Luk, 2010). Note that this aim is set up as a long term goal of scientific study for any particular (scientific) discipline.

The quality of scientific knowledge can be measured in terms of its reliability, its consistency and its accuracy. Good quality is judged by peer scientists when results of the research paper are reviewed (Figure 1), or when results in laboratory tests have acceptable level of precision. All three measures (e.g., reliability) of quality are relative to the current achieved level of performance (as reported in the literature). Accuracy is obviously important for scientific models to make predictions. Sometimes this performance is measured based on the notion of accuracy, for example precision and recall performance as the task may require the model to identify the desired items out of a collection. Accuracy may also be measured by precision (e.g., Rainville et al., 2005) when validating a scientific law, so accuracy may be measured in many different ways. Consistency can also be measured. Crude examples of measures of consistency include the number of important phenomena that are explained by the theory and the number of important anomalies that it cannot explain. The scientific knowledge is objective in the sense that its explicit form can be understood by other scientists unambiguously so that it can be tested by other (independent) scientists to assure its quality and its objectivity. The scientific knowledge should strive for generality. This avoids the simple accumulation of a large set of facts or experience as scientific knowledge (Kosso, 2007). Finally, the knowledge should be complete for the chosen domain of study. While scientific knowledge can be assessed in terms of logical necessity, reliability and accuracy, it does not imply that all scientific knowledge is automatically true, absolutely reliable and has no errors. Rather it is improved in terms of these aspects by scientific research. Often further research is needed because scientific knowledge of a general scientific discipline (e.g., physics) is usually incomplete, and proving this knowledge completeness may itself be a scientific endeavor.

According to Definition 1 in (Luk, 2010), scientific studies have at least one of the following three entities, i.e., theory, scientific model and experiment, so that they are distinguished from other types of studies. In this paper, such scientific studies are carried out by scientists (as in Figure 1) who are characterized as follows.

Definition 2: A scientist:

- (a) has or can acquire the (working) *scientific knowledge* of the domain; and
- (b) has or can acquire the enabling technical knowledge for her/him to conduct *scientific study*; and
- (c) uses methods and/or methodologies that can accomplish some or all aspects of the aim of scientific study;

In plain English, property (a) of a scientist requires her/him to know her/his subject area in terms of, for example, applying the theory to build models which are evaluated by experiments, and using the theory to explain phenomena in experiments. Property (b) of a scientist requires her/him to be able to conduct scientific studies. Otherwise, such a scientist cannot confirm, falsify, revise nor generate scientific knowledge. Property (c) ensures scientists use appropriate methods and/or methodologies (e.g., scientific method or research methodologies in sampling) that accomplish only some aspects (e.g., good quality and objectively-accessible knowledge) of the aim of scientific study because it is seldom possible to accomplish all aspects of this aim, especially obtaining complete scientific knowledge of a domain. In practice, scientists usually try to accomplish as many aspects of the aim of scientific study as possible in order to publish a paper demonstrating that superior research is done.

3.2. Principles and Assumptions

Our theory of scientific study does not have any physical laws (also known as empirical law, e.g., (Weber, 2004)) but it has a set of principles and assumptions, which are based on the aim of scientific study. There are also two types of principles: basic principles and guiding principles. A basic principle is one that is applicable to all scientific studies all the time. A guiding principle is one that scientists should follow when conducting scientific studies.

The basic principles in our theory of scientific study are stated as follows. The first basic principle applies only to scientific theories and does not apply to scientific models because such models are already based on evidence (e.g., prediction accuracy) from experiments, and because this principle ensures all scientific knowledge are testable as the aim of scientific study requires.

Basic Principle of Empiricism: *A scientific theory must be directly or indirectly based on evidence from experiments, which supports or potentially falsifies the theory.*

This basic principle means that the scientific theory can indirectly relate to experiments via the scientific model or have consequences in certain physical situations. At some course of time, a scientific theory must have stood either the direct test of falsification (Popper, 1959) or indirectly the validation of a scientific model that the scientific theory supports.

The next basic principle is useful for scientists to communicate and to reuse knowledge in scientific theories and scientific models.

Basic Principle of Theoretical Objectivity: *A scientific theory and its supported scientific models must be explicit so that they can be communicated to other scientists unambiguously and*

should be fully or partly formalized (e.g., mathematically or logically) for reasoning and testing inconsistencies.

This basic principle requires the scientific theory and its scientific models to be objectively accessible in mathematical forms or some formal forms so that inferences can be made. In this way, we may also check whether a theory is consistent with its model or not. In general, some scientific knowledge may not need to be consistent with each other (Aerts et al., 1999) because they (e.g., hypotheses) are "working knowledge" to facilitate scientific progress (through analyzing errors for example (Farrell and Hooker, 2009)). Some special type of scientific knowledge needs to be consistent and self-contained. This is specified by the following principle for scientists to follow rather than using definitions to categorize knowledge into scientific and non-scientific ones.

Basic Principle of Theoretical Consistency: *A scientific theory and its supported scientific models must not be inconsistent with each other and with themselves.*

The above basic principle focuses on consistency between scientific theory and its supported scientific models because the correct scientific theory and its supported scientific models will eventually be consistent, and because experiments do not need to be consistent with scientific models or theories as they are evaluated or tested by experiments. Consistency does not necessarily mean that the supported scientific model is logically implied by the scientific theory as in logical reduction (Aerts and Rohrlich, 1998). Instead, a scientific model that is supported by the scientific theory must have used some of the statements of that theory (e.g., its assumption, its principle or its law) so that the scientific model is related to the scientific theory. Effectively, we are forgoing the *mandatory requirement* of logical reduction and therefore logical atomism because it is difficult even for a mature scientific discipline (e.g., physics) to be complete, and so such a discipline lacks the complete

vocabulary and relationships to be able to fully specify the logical system for deduction. In the case of severe lack of scientific knowledge, some (e.g., (Damper, 2006)) may even consider thought experiments (as a way of deducing outcomes) to be harmful. Having said that such logical system is highly desirable to scientists, and in practice deduction sometimes is possible only for some focused area of study rather than the entire discipline.

In experiments, we require the following basic principle to hold in order to be objective and therefore unbiased:

Basic Principle of Objective Experiment: An *experiment* should not be intentionally biased to obtain a particular, favored outcome (e.g., favoring a particular model over other models, or a particular theory over other theories) by manipulating the experiment.

This principle is basic because if the scientist does not obey this principle, we want to claim that there may be misconduct in the experiment leading to some biased claim in the scientific study. To fulfill this principle, the experiments are usually documented to demonstrate that no bias is introduced in the set up and the procedure of the experiment, etc. This principle may be considered to be derived from the principle of honesty, but we want to state this principle because its specification is directly related to scientific study whereas the principle of honesty refers to general human conduct.

To connect physical situations of experiments with models, we assume that:

Basic Principle of Modeling Accuracy (Lower Bound): A scientific model should achieve statistically significantly better

prediction accuracy than random guesses using the appropriate minimal prior knowledge.

The principle does not say the desired accuracy of the scientific model since scientists try to establish the most accurate model of the physical situation through the social process of publication. There are, however, some bounds on the prediction accuracy of the model. The upper bound is determined by the precision of the instruments, the prediction limit of the physical laws, whether all factors are accounted for in the model, etc. The lower bound of the prediction accuracy is determined by the random model that guesses the outcome with the appropriate minimal prior knowledge since the scientific model that has the scientific knowledge needs to predict (statistically significantly) better than the random model that makes random guesses. Note that some scientific model cannot specify the absolute accuracy level. Instead, it may specify a relative accuracy level (e.g., 5%) better than random guesses or the baseline performance. In even more extreme cases, one may only be able to say that the scientific model accuracy is statistical significantly better than the baseline without even being able to specify how much the accuracy is better.

One part of the aim of scientific studies is to produce general knowledge which is guaranteed by the following:

Basic Principle of Generalization: The theory generalizes the applied (related) models which generalize the corresponding physical situations of the experiments.

Obviously, the model of the physical situation may leave out some details or factors, so the model may simplify and therefore generalize the physical situations in this way. In addition, the model may parameterize the physical situation so that the same model can be applied to many different but related

physical situations with different parameter values, thereby generalizing the physical situation by the model. A generalized model may generalize a number of different specific models together (e.g., by making a less restrictive assumption), so that the generalized model generalizes even more physical situations than the specific models. A theory generalizes a number of different models by applying the same principles or assumptions to the different models. In addition, a theory may be generalized by a general theory (e.g., general theory of relativity) by showing that the specific theory (e.g., Newtonian mechanics) is a special case or an approximation of the general theory under specific conditions, so that the general theory is applicable to even more models. In summary, the scientific knowledge generalizes the observations made in the experiments in different ways (e.g., parameterization, simplification, approximation, etc).

Guiding principles of our theory of scientific study are formulated for scientists to follow. Unlike basic principles, these principles are not true all the time because some scientists may not follow them as they are not aware of them. Our guiding principles encourage scientific studies to meet the aim of scientific study as stated in Section 3.1. These guiding principles include:

Guiding Principle of Reliability: Scientists should use methods to assess the reliability of their (working) *scientific knowledge* obtained by conducting *scientific studies*.

Guiding Principle of Investigation Objectivity: Scientists should enable other scientists to carry out the *scientific studies* for independent verification.

When formulating these principles, we do not explicitly require scientists to use reliable methods because a scientist who follows both the principle of reliability and objectivity will be unlikely to use an unreliable method,

especially when there is a social process to perform independent confirmation. The above two principles are guiding ones because some scientific study cannot replicate the experiment (e.g., meteoroid-impact hypothesis for the extinction of dinosaurs in historical science (Cleland, 2001)), so reliability measures estimated by repeated trials cannot be directly used and the historical event cannot be repeated for direct observation. Instead of replicating the experiment, historical scientists find pieces of evidence from diverse sources or of different character to establish whether a historical event has occurred or not. In this case, the reliability of some pieces of evidence can be obtained or estimated.

The final guiding principle is about the nature of physical laws and principles in (scientific) theories:

Guiding Principle of Immutable Laws and Principles:

Principles and (physical) laws in (scientific) theories should not change in time.

This is not a basic principle because some laws are found to be false (e.g., Moore's law) but they remain as laws as they are accepted by the community for some time.

Our theory of scientific study needs to assume that scientists are competent to carry out the scientific study. This is formulated as follows:

Assumption 1: Scientists are sufficiently trained to conduct or to be enabled to conduct *scientific studies*.

In practice, it is very difficult to find a scientist who knows all possible techniques, technologies or methodologies to conduct all the scientific studies in the subject. Therefore, this assumption specifies the minimum requirement

for a scientist to conduct the scientific study, i.e., the scientist should at least be able to learn the technique, to use the technology, to follow a methodology or to find qualified people to help (e.g., find engineers to help build a large cyclotron) in order to carry out the scientific study.

The next assumption is to safeguard the objectivity of scientific knowledge so that it is objectively accessible for independent verification, confirmation, validation, falsification, etc. as follows:

Assumption 2: Scientists express their work accurately in scientific communications.

The above is specified as an assumption rather than a principle because it is obvious and so general that it is applicable to any profession (rather than just science, e.g., engineers).

We want the scientists to be objective and impartial when they carry out the experiment. This is partly enforced by the basic principle of objective experiment and partly by the following assumption:

Assumption 3: Scientists strive to make unbiased, (adequately) accurate observations in experiments.

We have chosen to express this as an assumption rather than a basic principle because this is an obvious requirement to do any experiments, and because violating it may not constitute automatically that there is scientific misconduct depending on the extent of the bias or inaccuracies and the available information at the time.

According to the definition of scientists, they implicitly acknowledge that they adopt the aim of scientific study, so the following is assumed to be true.

Assumption 4: The domain of study using *scientific studies* (as activities) adopts the aim of scientific study.

The above assumes that scientific study looks for good quality, objective, general, testable and complete scientific knowledge. This ensures that investigators generalize their scientific knowledge and expand their knowledge to cover the entire chosen domain of study. It should be noted that the long-term aim of doing scientific studies is the accumulation of such knowledge. Therefore, it does not mean that scientists cannot engage in formative research, qualitative research or build conceptual models. Instead, a scientist can engage in any mode of scientific study that advances scientific knowledge in one or more aspects (e.g., advancing the reliability of the scientific knowledge or widening the scope of the scientific knowledge) and not necessarily all aspects whenever a paper is published.

Another assumption relates to the causality of events (Regopoulos, 1966). In here, it is stated as:

Assumption 5: In a scientific study, the phenomena observed in the physical situation have causes.

We do not restrict our causes to natural ones as in (Frankfort-Nachmias and Nachmias, 1996) because there may be causes in undiscovered dimensions. This assumption is needed because we want to study the observed phenomena, explaining them by the relevant scientific model or theory. Note that the phenomena may arise from complex processes with certain percentage of the processes causing the phenomena to be observed, so that the cause may only be a factor instead of the sole cause of the phenomena. Such causation may be too complex to be observed directly, so it may be

believed to be the case by scientists, explaining why causation is not formulated as a principle.

Following the previous assumption is the related assumption that:

Assumption 6: A phenomenon in a physical situation can be explained by some theory or model.

This assumption shows that the theory or model has explanatory power, which implies that (mature or developing) science has explanatory power. This is an assumption rather than a principle because some phenomenon may not have an explanation yet, but it is believed that the (future) theory or model will provide the explanation of the phenomenon.

In order to make generalizations across physical situations, we require the following assumption to hold:

Assumption 7: If similar or identical physical situations occur, then similar or identical physical situations will produce similar or identical (probabilistic) distributions of outcome, respectively.

Here, it is assumed that all relevant factors are considered for the physical situations to be considered similar or identical. Usually, when similar or identical physical situations do not produce similar or identical results, scientists will find some hidden factors to explain why these physical situations behave differently, instead of abandoning this assumption. In here, we specified that the distributions of the outcomes are similar or identical rather than the actual outcomes because this is more general as the distributions of outcomes cover the case of the actual outcomes. For some disciplines like Physics, it may be possible to control the experiment to get almost identical physical situations. However for some disciplines like

economics, it may be impossible to control the physical situations to be identical, in which case the physical situations can only be roughly similar to each other (e.g., economy). This assumption has an impact on the repeatability of the experiment as we expect similar physical situations result in similar behavior producing similar outcomes.

Finally, we do not have a principle of completeness since it is obvious that scientists try to establish a complete mastery of the subject which is implied by making assumption 4. In fact, many research works try to build a more complete picture of the subject by discovering new phenomena (e.g., black-body radiation). That is why some scientific field may appear to be only empirical reporting only about new phenomena, instead of improving theories or models, as the knowledge of the subject is highly incomplete.

4. An Initial Interaction Model of Scientific Study

Our initial theory of scientific study has a general abstract model of scientific study (Figure 2), but we cannot apply our theory to build this general model (as in scientific studies like physics as explained by Luk (2010)) because the model is not detailed enough. Therefore, this section extends the general model with more details into a more complete interaction model so that our theory of scientific study can apply its principles to this interaction model. Owing to the vast details of scientific knowledge in various disciplines, our interaction model of scientific study is delineated as a (common) template so that specific scientific disciplines may use the template to further extend their own logical model of scientific study for the particular discipline or to further extend their own domain ontology for the specific scientific discipline. In this way, we can have a common understanding (i.e., the common template) across different scientific disciplines so that we can use this common understanding to distinguish disciplines that are scientific.

As scientific study consists of theories, models and experiments, we will discuss the details of the theory, model and experiment entity clusters in each subsection. Then, we combine these entity clusters into one model or template, and show how our theory of scientific study is applied to the combined model (i.e., the common template). In the final subsection, we discuss the different types of knowledge in scientific study.

4.1 Theory Entity Cluster

A *theory entity cluster* is shown in Figure 3. This entity cluster corresponds to the theory entity of Figure 1 in (Luk, 2010). The theory entity cluster has a number of smaller entities including the aim, assumption, definition, fact, term, etc. This entity cluster has some special relationships, called *is-a* relationships. For example, the scientific theory entity has an *is-a* relationship with the theory cluster. The direction of this *is-a* relationship is from the scientific theory entity to the theory entity and this can be interpreted as "scientific theory is a theory". A mathematical theory is also a theory based on the interpretation of the *is-a* relationship between the mathematical theory entity and the theory entity. Connecting the scientific theory entity and the mathematical theory entity is a circle with the label "o". This label indicates that a theory can be a scientific theory or a mathematical theory or both. This is called the *overlapping* constraint in the EER notation. In the EER notation, the theory entity is called a *supertype* entity, and the scientific theory entity and the mathematical theory entity are its *subtype* entities. Characteristics (i.e., any attributes) of the supertype entity are inherited by its subtype entities. Another *is-a* relationship is between the principle entity, the basic principle entity and the guiding principle entity. In this case, there is a *disjoint* constraint between the participating entities in this *is-a* relationship, and this relationship can be read as "a principle is either a basic principle or a guiding principle but not both". This reflects the incompatibility between the basic principle which is supposed to be true all the time and the guiding principle

which should be true (although not all the time) in the context of scientific studies. In addition, there are full and partial specialization constraints. An example of partial specialization is the supertype theory which may not specialize into a scientific theory nor a mathematical theory. In EER notation, this is drawn with a single line from the supertype entity (i.e., theory) to the subtype entities (i.e., scientific theory and mathematical theory). An example of full specialization is the supertype “fact” which must specialize to a “base fact” or a “derived fact” (but not both because of the disjoint constraint). That is there does not exist any fact that is neither a base fact nor a derived fact. In EER notation, the full specialization constraint is drawn with two parallel lines from the supertype (i.e., fact) to its subtypes (i.e., base and derived fact).

In Figure 3, the mathematical theory includes both quantitative mathematical theory and logic-based mathematical theory. A scientific theory should have a mathematical theory to support it in order to avoid finding contradictions in the scientific theory. A theory can have any number of sub-theories or focused theories (which are just modeled as theories in Figure 3). A theory exists in some context that it is applicable. By inheritance, the scientific theories and the mathematical theories are applicable only in the contexts that are inherited from the corresponding supertype theory entity. There exist some theories that are neither scientific theory nor mathematical theory, and such a theory may be a qualitative theory that may eventually develop into a scientific theory after it is verified and quantified or formulated systematically.

applied scientific theory. However, the scientific models need not apply all the principles in the theory if these principles are not relevant to the scientific models. A (physical) *law* (e.g., Zipf law) is a generalization of observations in experiments. Typically, such law expresses quantitative relationship in an experiment, and such a relationship may be induced from data by curve fitting with a particular confidence level of statistical tests. Such a quantitative relationship may be considered as a generalization of data in the experiments. When these laws successfully explain many novel phenomena or when they are used to make many successful predictions in novel situations by constructing scientific models, they may become principles (as they are applied).

In Figure 3, the term entity refers to different kinds of objects (in the physical situation) or different kinds of properties. A term may be a scientific term, a common term or neither of them (e.g., a technical term or a mathematical term). However, some scientific disciplines (e.g., biology) may have to create many scientific terms to refer to the different kinds of objects (e.g., naming different species on Earth). Due to the vast quantities of new terms, they are organized into knowledge structures called ontologies (sometimes reduced to taxonomies, e.g., (Saracevic and Kantor, 1997)) which can group different terms together if they share similar characteristics so that they can be distinguished and reasoned about as a group of homogeneous objects. Such ontologies of specific scientific disciplines are being developed for scientific knowledge management (e.g., (Kingston, 2002)).

In Figure 3, the interaction between the fact entity and the term entity is borrowed from Figure 4-15 in (Hoffer et al., 2002, page 149) which is an EER diagram to express EER constructs. Specifically, fact is some relation between two or more terms. In our case, a fact may be an equation like “force” equals to “mass” times “acceleration”. This equation uses the equality relationship to associate the term *force* with the term *mass* and the term *acceleration*. In

Newtonian mechanics, these three terms are scientific terms, and this equation is a base fact which is one of its basic principles of mechanics. It is a base fact by virtue that it is accepted as a basic property of this scientific discourse; and it is not an axiom of a mathematical system. Having wrote that, axioms can be base facts because they generate the mathematics that is applied to quantify the scientific subject. Likewise, the derived facts may be corollaries, lemmas and theorems in the mathematical systems, which may be applied when constructing scientific models. Derived facts may be deduced from known universal statements logically.

Table 1 shows a summary of our theory in a table format that groups the components of our theory into entity types of the theory entity cluster in Figure 3. This illustrates how our theory fits into this theory entity cluster. Some term has one meaning as a common term and possibly another meaning as a scientific term (e.g., science in Table 1). Such terms are grouped under both scientific and common terms in Table 1. Many of these terms are defined in (Luk, 2010), so the definition entity in Table 1 has only two definitions as stated in this paper (i.e., Scientist and Enabling Technical Knowledge). An ontology of the terms is not constructed in this paper because we do not want to clutter up our diagrams. Our theory has an integrated interaction model in Section 4.4 which is a logical system that shows the application of the basic principle of theoretical objectivity. Although our theory and model are not scientific ones yet, they are expected to be consistent with each other as required by the basic principle of theoretical consistency. Our theory is verifiable (e.g., (Magnani, 1999)) so that it may be directly related to the experiment entity. For example, scientists may be asked in a survey-type experiment as to whether they hold the basic principles of our theory when they conduct scientific studies. In this way, the basic principle of empiricism may be applied to our theory.

| Entity Type | Details | Descriptions / Examples / Titles |
|--------------------|----------------------------------|--|
| Term | Scientific Term | Scientific Theory, Scientific Model, Computational Model, Computational Scientific Model, Simulation Model, Formative Scientific Study, Developing Scientific Study, Mature Scientific Study, Enabling Technical Knowledge |
| | Both Scientific and Common Terms | Science, Scientist, Experiment, Physical Situation, Theory, Model, Scientific Knowledge, Scientific Study, Research, Scientific Research |
| Aim | | Aim of Scientific Study |
| Assumption | 1 | Sufficiently Trained |
| | 2 | Accurate Communication |
| | 3 | Unbiased, Accurate Observations |
| | 4 | Adoption of the Aim of Scientific Study |
| | 5 | Causality of Phenomenon |
| | 6 | Explanatory Power |
| | 7 | No Magic |
| Principle | Basic Principle | Empiricism |
| | | Theoretical Objectivity |
| | | Theoretical Consistency |
| | | Objective Experiment |
| | | Modeling Accuracy (Lower Bound) |
| | Guiding Principle | Generalization |
| | | Reliability |
| | | Investigation Objectivity |
| | | Immutable Laws and Principles |
| Definition | 2 | Scientist |
| | 3 | Enabling Technical Knowledge |
| Theory | | Our Theory of Scientific Study in this paper |
| (Scientific) Model | | Integrated Interaction Model in Figure 7 |
| Context | | Scientific Study in Figure 1 |

Table 1: Components of our theory which are grouped according to the entity types of the theory entity cluster in Figure 3.

4.2 Scientific Model Entity Cluster

A *scientific model entity cluster* is shown in Figure 4. The scientific model entity is formulated by applying some principles in the theory entity. If the (physical) law is applicable, scientific models have to obey them. Scientific models also make some model-specific assumptions that are not made in the theoretical assumptions. For example, when a car is rolling down a slope, it is often assumed that there is no friction between the car and the slope. Such an assumption (sometimes known as an auxiliary assumption) is model-specific

because it is specific to this situation being modeled. In many practical cases, the model-specific assumptions are usually false but they are made to simplify the modeling of a physical situation. A scientific model can have scientific sub-models by adding more details to these scientific sub-models.

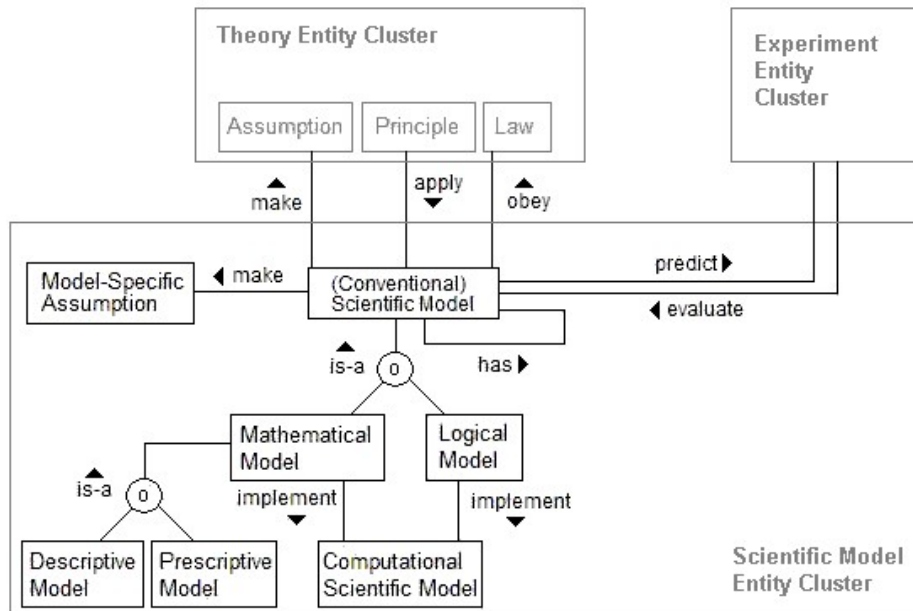


Figure 4: Details of the scientific model entity cluster that corresponds to the scientific model entity in the process model in Figure 1 of (Luk, 2010).

The general scientific model may make general predictions where as the scientific sub-models can make more accurate predictions in more limited situations. The possibility that scientific models can have scientific sub-models does not imply that micro-level scientific models must be the scientific sub-models of a macro-level scientific model. There is no logical necessity that reductionism (Nagel, 1974; Dieks and De Regt, 1998) automatically applies to the micro-level and macro-level scientific models, particularly when the instruments used to observe the macro-level and micro-level events are not the same or are unrelated. However, it is possible that some micro-level scientific models are scientific sub-models of a macro-level scientific model, depending on the particular subject of study.

A scientific model may be implemented in a computational scientific model as shown in Figure 4. In such a case, this scientific model is called a conventional scientific model in (Luk, 2010). Some computational scientific models are a generalized model of a number of conventional scientific models. In this case, there is a logical model that controls the operation of the different conventional scientific models in the generalized model. The existence of a mathematical model implies the existence of its logical version which is useful for hypothesis testing.

Building mathematical models (Hennig, 2009) has its own activities either as one part of the scientific studies or as one part of the engineering activities. These activities are summarized by Meyer in a flow chart in Figure 5. According to Meyer (1985), there are two types of mathematical models: descriptive models and prescriptive models. Descriptive models tell us how the objective of the study operates now and in the future, and scientific mathematical models are descriptive ones. Prescriptive models help us to choose the best way, and they are called optimization or normative models in engineering. Sometimes, prescriptive models are scientific models because the object of study is performing optimization. For example, an ant walking from one place to another is explained by a prescriptive model that optimizes the path of exploration. Therefore, a mathematical model can be both a descriptive and a prescriptive model. While Meyer (1985) provided some guidance on the selection of better mathematical models (e.g., accuracy, descriptive realism, precision, robustness, generality and fruitfulness), we restrain from a discussion here because this guidance may be task specific or domain specific, so we need to know the details of the specific application (which we lack) before we can discuss how the models are selected.

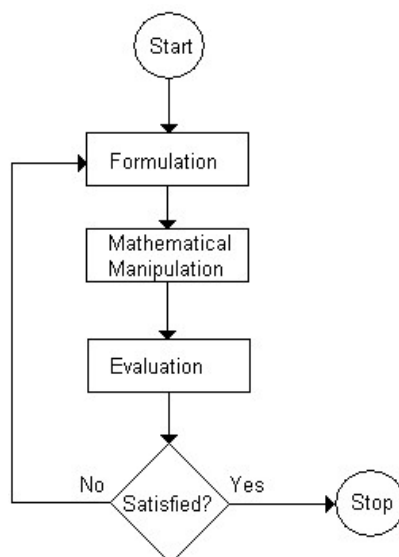


Figure 5: A flowchart diagram that illustrates the general mathematical modeling process after (Meyer, 1985).

4.3 Experiment Entity Cluster

An *experiment entity cluster* is shown in Figure 6. This cluster has an experiment entity which may be a quasi, controlled or natural experiment depending on the degree of control that the experiment has. The cluster also has a methodology entity that specifies what instruments to use and that determines how experiments should be set up and be conducted. In social sciences, there may not be any mechanical instruments. Instead, the methodology tells the social scientists how to conduct the experiment, and they may be exciting (e.g., by talking) and sensing (e.g., by listening) the physical situations directly. In this case, the instruments are just the scientist's mouth and ears. The scientist makes observations from the experiments. Such observations may be (textual) descriptions, or data points that may be presented visually (as a graph or a chart). If many experiments have the same or similar observations (which are data points), they may be generalized by induction to a general statement (e.g., an empirical, quantitative law) in a theory. Some observations arise from some phenomenon that may be explained by existing theories. Some unaccountable observations are called

anomalies, which cannot be explained by existing theories. The hypothesis entity in Figure 6 corresponds to abducted or formulated hypotheses that may be falsified or supported by evidence depending on the experimental findings.

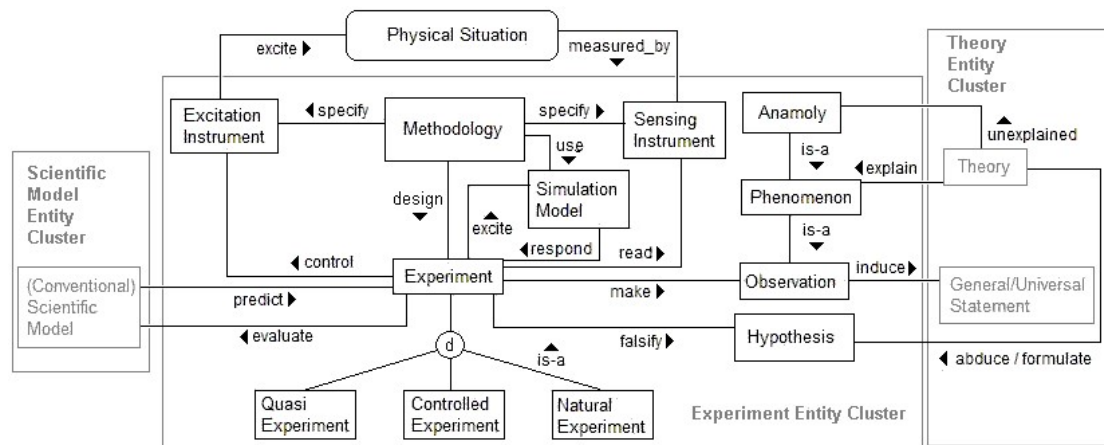


Figure 6: Details of the experiment entity cluster that corresponds to the experiment entity in the process model in Figure 1 of (Luk, 2010).

As shown in Figure 6, some experiments use simulation models as a replacement of the physical situation (Hartmann, 1996). For example, such models may generate data, the statistical properties (Humphreys, 1995) of which match those of the physical situations. In this way, it is possible to carry out experiments using the simulated data which do not match the statistical properties of the physical situations in order to observe novel phenomena in novel simulated situations. Some experiments use simulation models as one part of the experiment. For example, ELIZA (Wiezenbaum, 1966) is a program that simulates a human who responds to a human subject in a conversation. In this case, the physical situation is not replaced by the simulation model rather it includes the human subject and the computer program to generate the response in the experiment.

4.4 Combined Initial Interaction Model (Common Template)

Figure 7 integrates the interaction models in Figures 3, 4 and 6 in order to form a template for modeling scientific study of specific scientific disciplines. The interactions within individual entity clusters are quite complex and interactions across entity clusters are not trivial. This suggests that scientific studies are very complex processes. By organizing scientific study in terms of a theory and some logical models, we are able to capture a few common basic principles and theoretical assumptions from such complex processes.

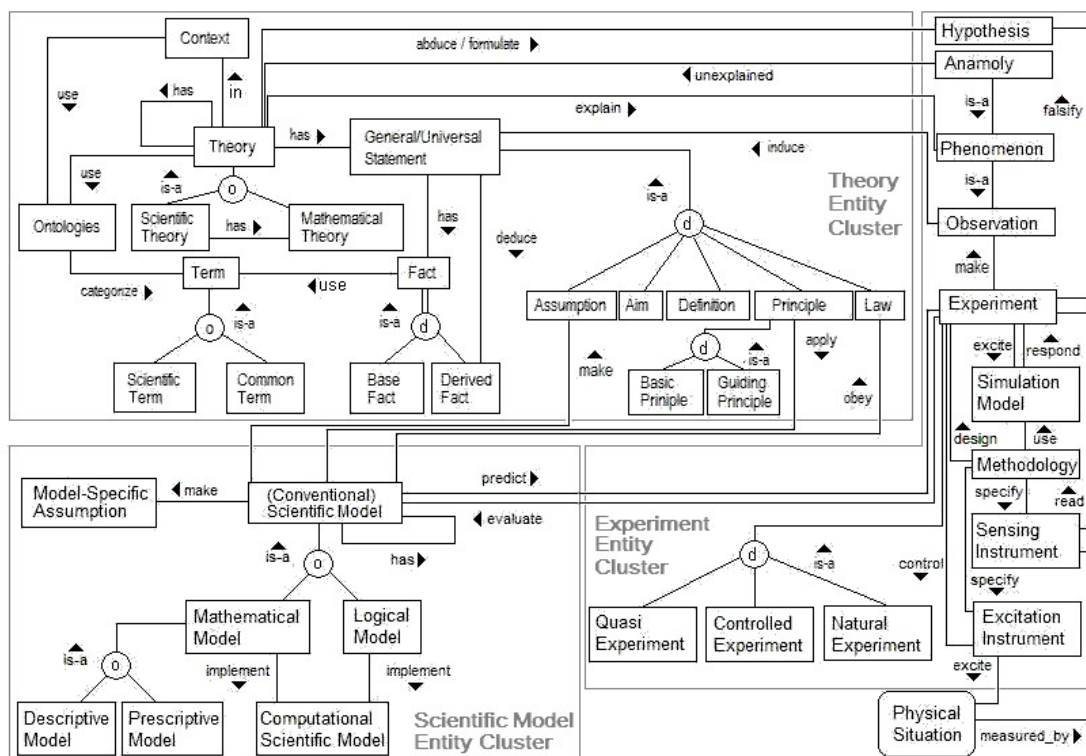


Figure 7: A detailed interaction model (i.e., the common template) of the process model in Figure 1 (Luk, 2010).

How is our theory of scientific study applied to the interaction model of Figure 7? First, our theory has assumptions, aims, definitions and principles which are present in the theory entity cluster in Figure 7. The principle of empiricism is applied to the scientific theory entity in Figure 7, which is required to be verified by experiment according to the principle. The principle of theoretical objectivity requires the interaction model and our theory to be partially formalized to facilitate reasoning and testing inconsistencies. The

principle of theoretical consistency requires that our theory and the interaction model are not inconsistent with each other; and so far no inconsistencies are found. The principle of objective experiment specifies the methodology entity of the experiment entity cluster of Figure 7 (i.e., the experiment methodology) not to favor any particular outcome by manipulating the experiment. The principle of reliability requires the experiment methodology of Figure 7 to assess the reliability of the scientific knowledge under investigation. The principle of investigation objectivity specifies that the experiment methodology of Figure 7 to report how the scientific study is carried out for independent verification. The principle of modeling accuracy highlights that the modeling accuracy of the scientific model entity of Figure 7 should be higher than that by random guessing. The principle of generalization requires our theory to generalize the interaction model as the principles are applied to the different interaction models for physics, for chemistry, etc. since Figure 7 is a generalized model combining the interaction models of physics, chemistry, etc. Assumption 3 requires the observation (entity) in Figure 7 to be made as much as possible in an unbiased and adequately accurate way. Assumption 4 assumes that the experiment methodology and the scientist hold the aim of scientific study when the methodology and scientist carry out the experiment. Assumption 5 requires that the phenomena observed in the physical situation (external entity) of Figure 7 to have causes which are modeled by the theory or the model under investigation. In summary, our theory of scientific study is highly interconnected with the interaction model, and the principles and assumptions are applied to specify how the entities of Figure 7 behave or what properties do the entities of Figure 7 possess. For completeness, note that assumptions 1 and 2 are applied to the scientist entity in Figure 1, so Figures 1 and 7 should be combined together to form a more complete model of scientific study that includes the contextual elements.

4.5 Knowledge

A scientist has both scientific knowledge and enabling technical knowledge as shown in Figures 1 and 2. In this subsection, the details about these two types of knowledge are discussed, and an interaction diagram shown in Figure 8 illustrates their structure and interaction. The diagram serves as a template rather than an exhaustive list of scientific and enabling technical knowledge because some enabling technical knowledge depends on the domain. A type of scientific knowledge typically uses more than one type of enabling technical knowledge. The remaining part of this subsection discusses scientific knowledge first and then the enabling technical knowledge.

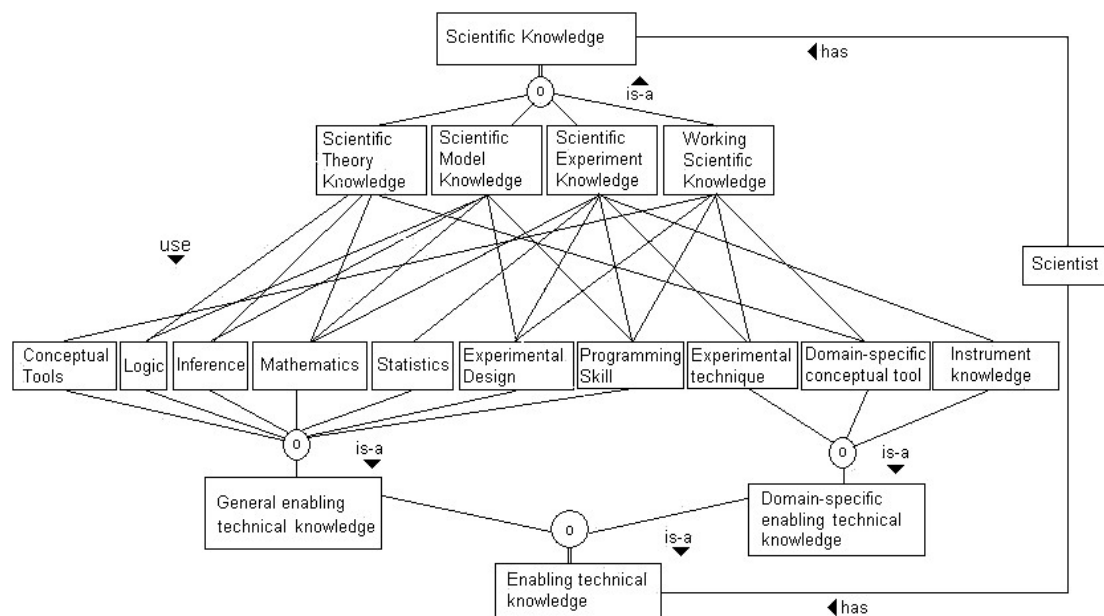


Figure 8: A scientific and enabling technical knowledge interaction diagram.

Scientific knowledge is already defined in (Luk, 2010). This section describes the different types of scientific knowledge. There are three general types of scientific knowledge: scientific theory knowledge, scientific model knowledge and scientific experiment knowledge. The scientific theory knowledge is a scientific theory or a set of scientific theories that are meant to be applied in a specific context. Each theory has a set of general/universal statements that are considered to be true or have not been falsified so far. Different types of

general statements include definitions, assumptions, principles and laws. Principles are divided into basic and guiding principles which were discussed in section 3.2 in the context of the theory of scientific study. (Physical) laws are empirical (mostly quantitative) relationships found in experiments and these laws may be stated as part of a scientific theory. A scientific theory requires that:

- (a) all the general statements and any sub-theories within the scientific theory should not be inconsistent with each other. This is required by the basic principle of theoretical consistency; and
- (b) these general statements and sub-theories have been tested experimentally. This is required by the basic principle of empiricism.

Scientific model knowledge is the technical knowledge of the scientific model specified in terms of mathematics and logic. For the computational model, the related algorithms to execute the computation quickly are actually enabling technical knowledge and not scientific model knowledge per se. Scientific model knowledge also includes assumptions. Some of these assumptions are shared with the theory but some of these assumptions are unique to the scientific models. Principles and laws are not considered to be part of the scientific model knowledge because they are applied to different scientific models, and because they already form one part of the scientific theory knowledge.

Scientific experiment knowledge is the knowledge about the experiments that enable them to be carried out for scientific studies. Such knowledge includes the aim of the experiment, experimental set up, procedure to carry out the experiment, instruments used, observations made during the experiment, etc. Such knowledge is needed to ensure the experiment can be repeated and the expected excitation and expected outcome are known for confirmation, comparison, validation and verification.

There exists one more type of scientific knowledge, which is not scientific theory, scientific model and scientific experiment knowledge. For example, there are theories, which are not ascended to the status of scientific theory because the theory is not fully consistent yet with some empirical evidence (i.e., anomalies), or because the theory contradicts some successful scientific models, or because the theory is a mathematical one that has not been fully developed. Other types of scientific knowledge bridge the gap between theories, scientific models and experiments. For example, hypothesis is a special type of scientific knowledge, which is an explanation of a phenomenon and is formulated for falsification. These types of knowledge are encompassed within the category of working scientific knowledge.

The knowledge boundaries between scientific theory, scientific model and scientific experiment may not be clear as these entities interact with each other. In particular, the general principles (e.g., $F = ma$), physical law, etc. belongs to the scientific theory, but their (partly) instantiated ones (e.g., $F = 0.5a$) may belong to the experiment entity. Similarly, the general principles, assumptions, etc. that are applicable to more than one type of situations belong to the theory, but their (partly) instantiated or parameterized ones specific to the model (as some of the variables may be derived from other formulae in the model) belong to the model entity. Likewise, the (partly) instantiated or parameterized formula specific to the model belongs to the model, and the corresponding instantiated formula or the formula with specific constants induced from data may belong to the experiment. In general, the knowledge boundaries would become clearer as we have all the details about the specific theory/model/experiment instances, and we have to decide the knowledge boundaries carefully case by case.

Apart from scientific knowledge, another type of knowledge facilitates scientists to conduct investigations, and it is called:

Definition 3: Enabling technical knowledge is a kind of knowledge that:

- (a) makes it possible to carry out the *scientific study* (e.g., building a particle accelerator); or
- (b) assures *scientific studies* (as activities) to produce knowledge of certain quality in terms of, for example, accuracy, reliability and consistency (e.g., deductions that ensure certainty of results or accepted experimental procedures); or
- (c) enables *scientific studies* (as activities) to assess the quality of the produced *scientific knowledge* (e.g., statistical significance tests);

Enabling technical knowledge can be divided into general and domain-specific. General enabling technical knowledge can be applied to different domains of study. For example, knowledge in logic and mathematics enable scientists to develop scientific models for different domains. Knowledge in inference (such as induction, deduction and abduction) helps scientists in valid reasoning independent of the domain of study. Knowledge in statistics enables scientists to assess the reliability of the experimental outcome. It should be noted that statistics is an application of probability theory, which is one branch of mathematics. Knowledge in experimental design enables scientists to design effective and efficient experiments to draw statistical conclusions. Knowledge of qualitative research methodology (e.g., interview and participant observation) enables scientists to make informed observation about human behavior, activity and organization. Knowledge of conceptual tools (e.g., schematic diagram) assists scientists to organize complex data and information for analysis. Knowledge about programming enables scientists to implement computational models, to gain knowledge and to qualitatively evaluate predictions from these computational models.

Domain-specific enabling technical knowledge is only useful in the domain of the study or a limited domain of the study. One such type of domain-specific enabling technical knowledge is the technical knowledge about building specialized instruments for experiments in the particular domain. For example, particle physicists need to study how to build particle accelerators and particle detectors in order to carry out experiments about the nature of particles. This example represents instrument knowledge that scientists possess for experimentation. Another type of domain-specific enabling technical knowledge is domain-specific conceptual tools. For example, Feynman diagrams are domain-specific conceptual tools for physicists to represent particle interactions. These diagrams are not used in other domains like biology or other studies of physics (e.g., thermodynamics).

Note that the enabling technical knowledge is subject to change as it may be further developed in other fields such as mathematics or logic. So, we do not expect that the enabling technical knowledge to be complete with all the important theorems discovered. Instead, we expect that the foundation (e.g., axioms) of the enabling technical knowledge should be built in order to ensure the proper application of the enabling technical knowledge to the scientific discipline. Important theorems may be discovered later which may drive further development of the scientific discipline. As enabling technical knowledge is under constant update with new results, it is hard to draw a fixed line between working and established enabling technical knowledge. Coupled with the fact that enabling technical knowledge is separate from the scientific knowledge where the new updates of enabling technical knowledge may not be relevant to the scientific knowledge, we draw only the enabling technical knowledge entity in our figures, implicitly assuming that it is under constant update as a field.

It is unrealistic to expect a scientist to know all the enabling technical knowledge. However, they are expected to be able to learn the necessary enabling technical knowledge if they are required by their particular scientific studies. Such requirement of scientist's capability is stated in assumption 1. Having said that, there is some core enabling technical knowledge that scientists must have. Such technical knowledge enables scientists to manage scientific theories, scientific models and experiments in general. Therefore, scientists are expected to have background training in logic, inference, mathematics, experimental design and statistics because:

- (a) logic and inference enable scientists to manage most theories;
- (b) logic, inference and mathematics enable scientists to manage most scientific models;
- (c) experimental design enables scientists to manage most experiments;
- (d) statistics enables scientists to make assessment of many different types of scientific knowledge.

The scope of the background training depends on the extent to which logic, inference, mathematics, experimental design and statistics are used in the particular (scientific) discipline.

5. Types of Research and Scientific Study

Research and scientific study are not always synonymous (Luk, 2010). First, the aim of research is to make some advancement in a chosen field of study. Therefore, research focuses on any aspect that deals with gaining novel knowledge. The novelty in research may be using a new approach to solve an old problem, an identification of a new problem that needs attention, etc. By contrast, scientific study is not necessarily concerned with generating new scientific knowledge. For example, government scientists may be routinely engaged in known ways of doing scientific studies for tracking the outbreak of epidemics in order to safeguard world health. In this case, the scientists may publish government reports to describe the case instead of publishing a

journal paper or a conference paper to report any novel work done in the subject.

Second, research may sacrifice reliability for novelty. In formative research, the researchers may wish to explore the "landscape" of the research topic by using explorative research techniques. For example, in engineering, prototypes are designed to test the feasibility of an approach to a problem. This prototype is used as a vehicle by the engineer and the user to explore the potential problems. Another example is the use of qualitative research methodology to gather opinions and views about certain social subjects before a quantitative survey is designed and administered. Such formative research may be found in formative scientific studies (Luk, 2010).

When the research matures as in mature scientific studies (Luk, 2010), concern over reliability, accuracy and consistency become more important than novelty. In this case, mature research may become scientific study in which some scientific theory is established, a host of scientific models are set up and a host of experiments are well known. It is not uncommon for some discipline to evaluate formative research as mature research. On the other hand, results in formative research need independent confirmation to ensure the reliability and objectivity of the results.

6. Developing a Scientific Theory of Scientific Study

Our theory of scientific study is not a scientific theory because it has not been tested empirically. Instead, it is based on some cases and discussed issues in philosophy of science. It is possible to launch a scientific study to establish a scientific theory of scientific study. Such a scientific study belongs to social science because the object of study is the behavior of, the practice of, the knowledge of and the organization of scientists.

Before launching a scientific study for developing its scientific theory, fundamental issues (e.g., knowledge boundaries of entity clusters) of our initial theory need to be discovered and debated in order to better guide such a scientific study. Even though these fundamental issues may not be possible to resolve, the awareness of these issues is important to scientists who carry out such a scientific study. Such awareness can guard against ignorant biases or misinterpretation of controversial observations. For instance, the work by Luk (2010) has integrated some issues of philosophy of science. By using his process model to develop our theory, it better informs us about the philosophical issues that are present in our theory and models of scientific study.

To establish a scientific theory of scientific study, it is necessary to establish what basic principles are held by scientists, what guiding principles are followed by scientists, and whether the interaction model fully describes the scientific study. The development of such a theory can begin with developing a fuller interaction model of different subjects (e.g., physics, biology, chemistry, psychology, etc) and merge these models of scientific study of particular domains together by discovering their commonalities and differences. Such a strategy to build a combined interaction model or EER diagram is known as the bottom-up approach. However, this approach runs the risk of developing fragmented models using different terminologies, and it may also bury the data models in the nitty-gritty of scientific activities that may lose sight of the philosophical issues that need to be addressed in the data models.

An alternative to the bottom-up approach is ours which develops a template (i.e., Figure 7) that serves as a unifying theme to merge the data models developed for individual scientific disciplines. Such an approach to building a combined EER diagram is known as the hybrid approach. This approach helps to maintain the coherence of the general knowledge structure which is

designed to take account of the philosophical concerns in the past (Luk, 2010), while encouraging the data models to use some common terminology. For building ontologies, our common template corresponds to the intermediate level ontology of Soldatova and King (2006), which can be used to develop domain ontologies for the different scientific disciplines. So, our template can be used for building EER diagrams or (domain) ontologies.

7. Related Work

Our theory of scientific study is different from the scientific method (Pierce, 1878; Weston, 1987) or the PEL (i.e., Presupposition, Evidence and Logic) model of scientific inquiry (Gauch, 2003). Apart from being more detailed, our theory involves the social learning process of scientific study, as well as enabling technical knowledge which is absent from the scientific method. Moreover, our theory organizes knowledge about scientific study in terms of an aim, a few principles, seven assumptions and a group of abstract models from general to specific ones. Such an organization of knowledge about scientific study has not been adopted by others nor those that develop systems for scientific knowledge management (e.g., (Hars, 2001)).

Gauch (2003) made four bold claims about the qualities of science. These qualities are rationality, truth, objectivity and realism. Truth and realism are desirable properties of statements in theories and in scientific models, respectively, of the process model of Luk (2010). Rationality is supported by the basic principle of theoretical consistency, and objectivity is supported by the basic principle of theoretical objectivity and the guiding principle of investigation objectivity of our theory. Whether these four claims hold for the scientific knowledge of a particular discipline depends on the extent to which the scientists follow the (basic and guiding) principles and whether their theories and scientific models are true and accurate, respectively.

Our theory inherits the terminology and properties of the process model by Luk (2010). The general model of our theory differs from his process model by extending it to model the social learning process of scientific study and to model its knowledge elements (e.g., scientific model) as entity clusters. In addition, our theory includes its aims, principles and assumptions which are absent in (Luk, 2010).

Our theory is different from the theory of idealization (Liu, 2004) in that his theory regards both models and theories as idealizations (Nowak, 1972; McMullin, 1985) whereas ours regards principles and laws in theories as the underlying true relationship found in physical situations and scientific models as approximating the physical situations (Niiniluoto, 1987; Marquis, 1991). Our theory is organized like a scientific one as ours has principles and a general (interaction) model of scientific study whereas the idealization theory is not organized as such. Our theory also includes the social process of publication (Figure 1) that is absent in the idealization theory. Such a social process is important because it encourages scientists to make their knowledge and its quality objectively accessible, thereby serving the aim of scientific study.

Our interaction model of scientific study is at a higher level than EER models or ontologies because the interaction model does not have the specific details about the axioms/rules, *part-of* relations, etc., so that our interaction model can be converted to EER models or ontologies by adding more specific information to it. We are unaware of any model of scientific study even though there is a proposed ontology for scientific experiments by Soldatova and King (2006). Apart from our interaction model being at a higher level than the intermediate level ontology of Soldatova and King (2006), our model explores horizontally the different aspects of scientific studies instead of vertically to all the low-level detailed concepts as explored by Soldatova and King (2006). In addition, Soldatova and King (2006) only provide an ontology

for experiments whereas we provide our model for scientific study which includes experiments, so our model is more general and complete than the ontology by Soldatova and King (2006).

8. Conclusion

We have developed a theory of scientific study as a social learning process (Figure 1) of scientists creating, revising, applying, monitoring (e.g., confirming) and disseminating (working) scientific knowledge. This theory is not a scientific theory yet, because it lacks detailed quantification to support the construction of a scientific model. However, it has its aim, a set of principles and assumptions which scientists are expected to follow or to acknowledge implicitly. Our theory also shows that a scientist does not just possess scientific knowledge but also enabling technical knowledge (that has often gone unnoticed). Our theory has a general interaction model (Figure 7), and we showed how our theory is applied to this interaction model. This interaction model (with EER constructs) serves as a template for those who want to develop specialized EER diagrams for knowledge management of specific scientific disciplines or to develop domain ontologies specific for the different scientific disciplines.

Acknowledgement

I would like to thank Dr. Edward Dang and the anonymous reviewers for the constructive comments.

References

Aerts, D., Brokekaert, J. and Smets, S. (1999) Inconsistencies in constituent theories or world views: quantum mechanical examples, *Foundations of Science* 3(2): 313-340.

- Aerts, D. and Rohrlich, F. (1998) Reduction, *Foundations of Science* 3(1): 27-35.
- Brodaric, B. and Gahegan, M. (2006) Representing geoscientific knowledge in cyberinfrastructure: challenges, approaches and implementations. Sinha, A.K. (Ed.) *Geoinformatics: Data to Knowledge*, pp. 1-20, Geological Society of America, Inc., Colorado, USA.
- Cleland, C.E. (2001) Historical science, experimental science and the scientific method, *Geology* 29(11): 987-990.
- Damper, R.I. (2006) Thought experiments can be harmful, *The Pantaneto Forum* 22.
- De Roure, D., Jennings, N.R. and Shadbolt, N.R. (2003) The semantic grid: a future e-science infrastructure. F. Berman, G. Fox and T. Hey (Eds.) *Grid Computing - Making the Global Infrastructure a Reality*, 437-470, Wiley, England.
- Dieks, D. and De Regt, H.W. (1998) Reduction and understanding, *Foundations of Science* 3(1): 173-188.
- Farrell, R.P. and Hooker, C.A. (2009) Error, error-statistics and self-directed anticipative learning, *Foundations of Science* 14(4): 249-271.
- Frankfort-Nachmias, C. and Nachmias, D. (1996) *Research methods in the social sciences*, Arnold Press, London.
- Gauch, H.C. (2003) *Scientific Method in Practice*, Cambridge University Press, Cambridge.
- Hars, A. (2001) Designing scientific knowledge infrastructures: the contribution of epistemology, *Information Systems Frontiers* 3(1): 63-71.
- Hartmann, S. (1996) The world as a process: simulations in the natural and social sciences. Hegselmann R., Mueller, U. and Troitzsch, K.G. (Eds) *Modeling and Simulation in the Social Sciences from the Philosophy of Science Point of View*, Kluwer Academic Press, 77-100.
- Hennig, C. (2010) Mathematical models and reality: a constructivist perspective, *Foundations of Science* 15(1): 29-48.
- Hoffer, J.A., Prescott, M.B. and McFadden, F.R. (2002), *Modern Database Management*, Prentice-Hall International, New Jersey.
- Humphreys, P. (1995) Computational empiricism, *Foundations of Science* 1(1): 119-130.

- Kingston, J. (2002) Merging top level ontologies for scientific knowledge management, *Proceedings of the AAAI Workshop on Ontologies and the Semantic Web*, Edmonton, Canada.
- Kosso, P. (2007) Scientific understanding, *Foundations of Science* 12(2): 173-188.
- Lakatos, I. (1977) *The Methodology of Scientific Research Programmes*, edited by J. Worrall and G. Currie, Cambridge University Press, New York.
- Latour, B. (1987) *Science in action: how to follow scientists and engineers through society*. Harvard University Press, Massachusetts.
- Luk, R.W.P. (2010) Understanding scientific study via process modeling. *Foundations of Science* 15(1): 49-78.
- Liu, C. (2004) Laws and models in a theory of idealization, *Syntheses* 138(3): 363-385.
- Ludäscher, M., Lin, K., Bowers, S., Jaeger-Frank, E., Brodaric, B. and Baru, C. (2006) Managing scientific data: from data integration to scientific workflows. Sinha, A.K. (Ed.) *Geoinformatics: Data to Knowledge*, pp. 109-130, Geological Society of America, Inc., Colorado, USA.
- Magnani, L. (1999) Withdrawing unfalsifiable hypothesis, *Foundations of Science* 4(2): 257-268.
- Marquis, J.-P. (1991) Approximations and truth space, *Journal of Philosophical Logic* 20(4): 375-401.
- McMullin, E. (1985) Galilean idealization, *Studies in the History and Philosophy of Science* 16, 247–273.
- Meyer, W.J. (1985) *Concepts of Mathematical Modeling*, McGraw Hill, New York.
- Nagel, T. (1974) What is it like to be a bat? *Philosophical Review* 83(4): 435-450.
- Niiniluoto, I. (1987) *Truthlikeness*, Reidel, Dordrecht.
- Nowak, L. (1972) Laws of science, theories, measurement, *Philosophy of Science*, 39(4): 533–548.
- Pierce, C.S. (1878) Deduction, induction and abduction, *Popular Science Monthly*, 13:470-782.
- Popper, K.R. (1959) *The Logic of Scientific Discovery*, Hutchinson, London.
- Pratten, S. (2007) The scope of ontological theorising, *Foundations of Science* 12(3): 235-256.

- Rainville, S., Thompson, J. K., Myers, E. G., Brown, J. M., Dewey, M. S., Kessler, E. G., Jr., Deslattes, R. D., Börner, H. G., Jentschel, M., Mutti, P., & Pritchard, D. E. (2005). A direct test of $E = mc^2$. *Nature*, 438(22), 1096–1097.
- Regopoulos, M. (1966) The principle of causation as a basis of scientific method. *Management Science* 12(8): C135-C139.
- Soldatova, L.N. and King, R.D. (2006) An ontology of scientific experiments, *Journal of the Royal Society Interface* 3(11): 795-803.
- Saracevic, T. and Kantor, P.B. (1997) Studying the value of library and information services. Part II. Methodology and taxonomy, *Journal of the American Society for Information Science and Technology* 48(6): 543-563.
- Silberschatz, A., Korth, H.F. and Sudarshan, S. (2005) *Database System Concepts*, 5th edition, McGraw Hill, New York.
- Weber, E. (2004) Rationality in the discovery of empirical law, *Foundations of Science* 4(3): 356-368.
- Weston, T. (1987) Approximate truth, *Journal of Philosophical Logic* 16(2): 203-227.
- Wiezenbaum, J. (1966) ELIZA – A computer program for the study of natural language communication between man and machine, *Communications of the ACM* 9(1): 36-45.

Biography: Robert Luk received a B.Sc. degree, a Dip. Eng. and a Ph.D. degree from the School of Electronics and Computer Science, University of Southampton, and M.Sc. degree from the Department of Psychology, University of Warwick. He is a visiting research scholar at the Center for Intelligent Information Retrieval, University of Massachusetts, Amherst, USA. He has published over 100 refereed publications (over 40 journal papers), is a PC member of various conferences (e.g., ACM SIGIR) and a reviewer of various journals (e.g., ACM Trans. Asian Language Information Processing and IEEE Trans. Systems. Man and Cybernetic). He is an associate professor of the Department of Computing, The Hong Kong Polytechnic University.